

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-091371

(43)Date of publication of application : 04.04.1997

(51)Int.CI

G06K 9/03

G06T 11/60

(21)Application number : 07-243166

(71)Applicant : NIPPON TELEGR & TELEPH CORP <NTT>

(22)Date of filing : 21.09.1995

(72)Inventor : YUHITO MITSUHIRO

INADA YOSHIAKI

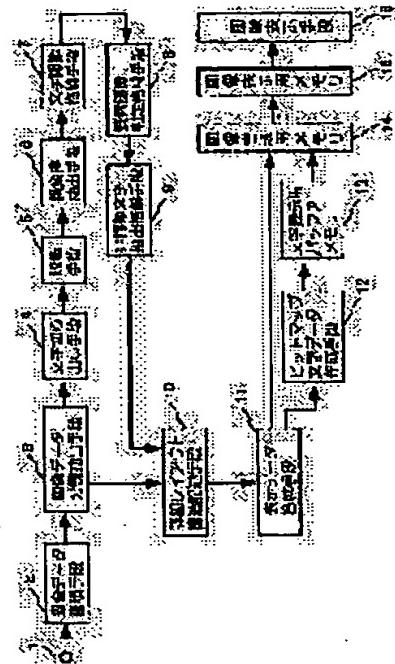
HORIGOME KAZUYUKI

(54) CHARACTER DISPLAY DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To provide the character display device which can display characters together with images of illustrations, etc., without correcting characters which were not recognized or erroneously recognized by a character recognizing means.

SOLUTION: The character display device is equipped with an element image deciding and storing means 8 which discriminates between a non-character image that does not consist of characters at all and a character element image that consists of characters entirely according to the recognition result of a recognizing means 5 and stores image data of the element image as a non-character image part when the image is the non-character element image, an unrecognized character extracting and storing means 9 which extracts image data of a character that the character recognizing means can not recognize and the position of the data in the image and stores them as an unrecognized character file, and a display data composing means 11 which puts bit map data for display generated by a bit map data generating means 12 and the image data stored in the non-character image part file together to generate data for display.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

文字認識結果を読み出し、表示用の文字のビットマップデータを作成するビットマップデータ作成手段と、前記ビットマップデータを表示用の字母に変換し、表示手段に表示させる画像表示手段とを組みた文字表示装置であって、前記粗雑手段の結果得られた各要素画像が全て文字以外のイメージ画像データから構成されている非文字要素四角か、全て文字のイメージ画像データから構成されている文字要素四角かを判定し、非文字要素四角であった場合には該当する要素画像のイメージ画像データを非文字四角部ファイルとして格納する要素画像判定手段と、文字要素画像内で前記文字既存手段が認識できない文字のイメージ画像データと該イメージ画像データの画像上位位置とを抽出し、非認識文字ファイルとして格納する非認識文字抽出粗雑手段と、前記ビットマップデータ作成手段が作成した表示用のビットマップデータと非認識文字ファイルおよび非文字画像部ファイルに格納されるイメージ画像データとを合成し、表示用のビットマップデータを作成する表示データ作成手段とを具備する。

[0013]前述した手段によれば、まず、画像データ分割抽出手段がイメージ画像データを要素画像に分割して各要素画像の全体画像中の位置を抽出し、次に、文字切り出し手段が前記要素画像内の文字画像を1文字づつ切り出す。

[0014]次に、認識手段が1文字づつ切り出した文字を該当する文字コードに対応させる文字認識を行い、文字認識の後処理として、誤差検出手段が文字認識で

【0015】 読み絞り所の検定を行なうと、文字認識手段は、前記認識手段の結果から各要素間の関係を考慮して、該要素が該文書の何處に該当するかを決定する。すなはち、該要素が該文書の何處に該当するかを決定する。

データから構成されている非文字要素画像から、全て文字のイメージ画像データから構成されている文字要素画像を判定し、非文字要素画像であった場合は該当する要素画像のイメージ画像データを非文字要素部ファイルとして格納する。

[0017] 文字認知を行った要素画像内に認識できないう文書があった場合には、該要素文字抽出手段が認

【0018】このとき、オペレータから表示指示がある場合、まず、ビットマップ文字データ生成手段が文字認識結果を読み出し、該表示用のヒットマップデータを作成する。

【0019】次に、表示データ合成手段が前記表示用示し手段によってビットマップデータと非認識文字ファイル(オルより非文字)の間に構成されるイメージ画像データとを合成し、表示用の合成ビットマップデータを作成して画面表示手段に送る。

100201 画面表示手段は、合成ビットマップデータを表示用の信号に変換し、表示手段に表示させる。

100211 このように、認識手段が文字として正確に認識できた文字については文字コードを格納し、一方、文字として認識できない、あるいは、誤認識してしまった文字についてはその文字のイメージ画像データを非認識手段から直接読み出した文字ファイルとして格納しておき、ナベレータから表示用示し手段が前記表示用示し手段によってビットマップデータと、非認識文字ファイルとを位置情報に基づいて組合することにより、文字として認識できない文字および、文字として認識することができない文字およびひび割れになってしまった文字をオペレータが修正することができます。

【0022】「発明の実施形態」以下、本発明について、発明の実施形態（実施例）とともに図面を参照して詳細に説明する。

【0023】なお、発明の実施形態を説明するものは同一符号を付け、図において、同一機能を有するものは同一符号を付ける。その限り返しの説明は省略する。

【0024】図1は本発明の一実施形態の文字表示装置である。

リ、1・5は画像表示用メモリ、1・6は画像表示手段である。

[0025] 図1において、イメージ画像データ入力端子1はイメージ画像データを本文端形態の文字表示手段に読み込むための読み込み端子であり、図示しないイメージキャラクターあるいはイメージ画像データが格納されられる。図示しない外部記憶装置が接続される。

[0026] 画像データ蓄積手段2は、イメージ画像データ入力端子1から読み込んだイメージ画像データをデータ入力端子1から読み込んだイメージ画像データから読み込まれたデータを蓄積する。また、本実施例においては、主メモリ上に確保される。

[0027] 画像データ分離抽出手段3は、画像データ蓄積手段2に蓄積されたイメージ画像データから読み込まれたデータを分離する。また、各画面表示手段1・2・3・4・5・6に對象となる全画面箇中で、画像を構成する単位プロトクルである複数の要素画像の位置を抽出し、各画面表示手段1・2・3・4・5・6に該位置を示す位置情報を付与する。

背号を示すものであるかを示す識別子と、前記要索画像の位置などを1つのファイルに格納する。

【0028】文字切り出し手段4は、前記画像データ一分割抽出手段3により読み込みごとに抽出(分割)された要素画像を、公知の技術を用いて、1文字単位のイメージ画像(文字画像)として切り出す。

【0029】認識手段5は、文字切り出し手段4で切り出された文字画像を1文字づつ順番に読み込み、たとえば、バターンマッチング法により、1文字づつに分割されれた文字画像と、予め定められたテンプレートとを比較・照合することにより文字画像に該当する文字を特定し、文字コードに変換する、いわゆる、文字認識を行う。

【0030】検索検査抽出手段6は、公知の技術を用いており、認識手段5で得られた文章を予め用意してある知識辞書等の辞書を基に認知結果の検証を行うことにより、誤認箇所の検定を行う。

【0031】文字認識結果手段7は、認識手段5で認識された文字の文字コードを図示しないメモリに格納する。

【0032】文字表示用データ一3は、ビューワー上に表示用ビューワー上に表示用データを作成した表示用ビューワー上に表示用データを作成した表示用データを記憶するためのメモリである。

【0033】画像合成用メモリ14は、表示データ合マップデータを記憶するためのメモリである。

【0034】画像合成手段15は、表示データ合マップデータを合成し、一時的に記憶しておくメモリである。

【0035】画像表示用メモリ16は、図示しない表示装置に表示させるためのビューワーを格納するためのメモリである。

【0036】画像表示手段17は、画像表示用メモリ15に格納されたビューワーデータを順次読み出し、表示しない表示装置に画像を表示させるためのアナログ信号に変換する。

【0037】【0038】【0039】【0040】【0041】図2は本実施形態の文字表示装置の動作説明するための入力画像を示す図であり、文字と図版含む。

[0032] 要緊面像判定結果手段 8は、公知の技術を用いており、各要緊面像内で文字として認識された部分がわかるか否かを調べ、たとえば、要緊面像内に文字として認識された部分が含まれていない場合は、その要緊面像は文字を含まない非文字要緊面像、すなわち、図版と判断し、この要緊面像の位置を抽出してその位置ヒメージ面像データとを面像ファイル（非認知文字符ファイル）として格納する。

[0033] 要緊面像判定結果手段 8は、公知の技術を用いており、各要緊面像内に文字として認識された部分がわかるか否かを調べ、たとえば、要緊面像内に文字として認識された部分が含まれている場合は、その要緊面像は背景面像を示しており、図像データ分析判断手段3は図3に示すように分割した面像を背景面像をファイル名とする背景面像ファイルとして、図示しない外に要緊面像1～8を示す。

[0042] 図3は図2に示す入力面像を本実施形態により、文字符および図版からなる要素に分割した結果を示しており、301～308は取扱いに要緊面像1～8を示す。

[0043] 図3において、たとえば、要緊面像8（08）は背景面像を示しており、図像データ分析判断手段3は図3に示すように分割した面像を背景面像をファイル名とする背景面像ファイルとして、図示しない外

【003-3】非構造文字抽出結果手段9は、要絆画像判定定格納手段8で文字を含む要絆画像であると判断された文字の要絆画像の内、認識手段5で認識された文字と、認識できなかった文字の文字画像と、それぞれのレイアウトでの位置情報をファイルに記述し、たとえば、図示しない外部記憶装置に格納し、保存する。

【003-4】該当レイアウト構造記述手段10は、画像データ公開抽出手段5が抽出した要絆画像の入力画像データから該当する該当レイアウト構造を抽出する。

四十七

【図9】文字認識結果の動作を説明するための図で

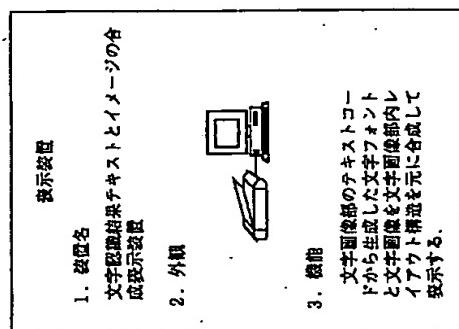
文獻研究 · 2006年第1期 · 11

ある。【図10】図9に示す超微結果をファイルに保存したときの記述例を示す図である。

手稿 3 …圖像子一名分類列出手稿 4 …文字切り出上 19

121

2



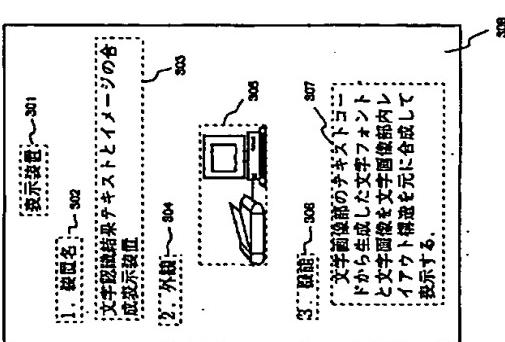
四

4



16

6



三

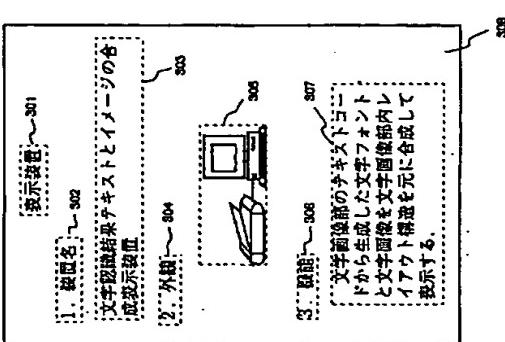
特開平9-91371

11

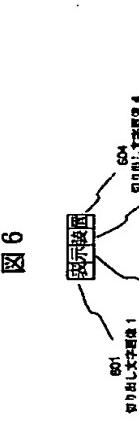
手段、5…認識手段、6…假想換抽出手段、7…文字題目手段、8…要素圖面判定分析手段、9…非認識文書抽出手段、10…詳細レイアウト構造記述手段、11…表示データ合成手段、12…ビットマップ文字データ作成手段、13…文字表示用バシフメモリ、14…圖像合成用メモリ、15…圖像表示用メモリ、16…圖像顯示手段、301～308…要素圖面1～8、601～604…切り出し文字圖面1～4、701…切り出し文字圖面5、702…切り出し文字圖面6。

31

62

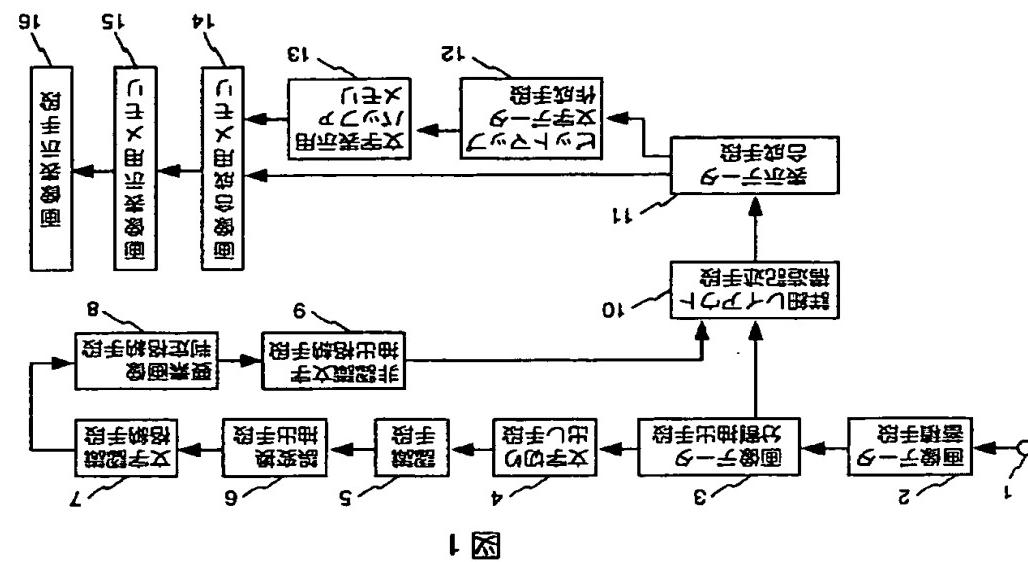


61



16

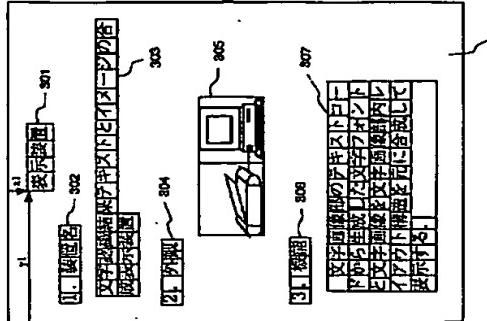
6



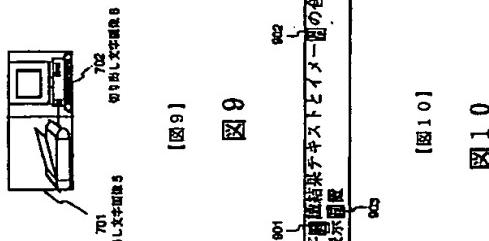
11

1

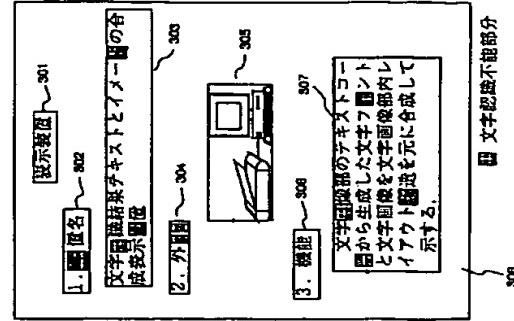
51



11



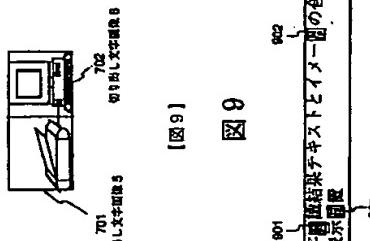
18



四一



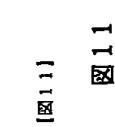
7



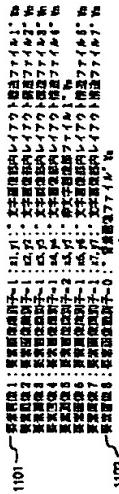
101



101



文字圖書



四、社會問題不均衡